



Designing Learning Environments for Critical Thinking: Examining Effective Instructional Approaches

Dawit Tibebe Tiruneh¹ · Mieke De Cock² ·
Jan Elen³

Received: 20 December 2016 / Accepted: 25 May 2017

© The Author(s) 2017. This article is an open access publication

Abstract Fostering the development of students' critical thinking (CT) is regarded as an essential outcome of higher education. However, despite the large body of research on this topic, there has been little consensus on how educators best support the development of CT. In view of some of the controversies surrounding the teaching of CT skills in higher education, this study examined the effects of embedding CT instruction systematically in domain-specific courses (*Immersion* vs. *Infusion*) on the acquisition of domain-specific and domain-general CT skills and course achievement. First-year university students ($N = 143$) enrolled in an introductory physics course were assigned to one of three instructional conditions: Immersion, Infusion, and control. The Immersion and Infusion conditions followed lessons designed systematically based on the First Principles of Instruction model, whereas the control condition followed a regular instruction. Results showed that participants in the Immersion and Infusion conditions significantly outperformed those in the control condition on domain-specific CT proficiency and course achievement. However, neither the Immersion nor the Infusion condition was helpful in fostering the acquisition of domain-general CT skills. The findings generally demonstrated that embedding CT instruction systematically in

✉ Dawit Tibebe Tiruneh
dawittibebe@yahoo.com

Mieke De Cock
Mieke.Decock@kuleuven.be

Jan Elen
Jan.Elen@kuleuven.be

¹ Department of Teacher Education and Curriculum Studies, College of Education and Behavioral Sciences, Bahir Dar University, PO. Box 79, Bahir Dar, Ethiopia

² Department of Physics and Astronomy & LESEC, KU Leuven, Celestijnenlaan 200C, Box 2406, 3001 Leuven, Belgium

³ Centre for Instructional Psychology and Technology, KU Leuven, Dekenstraat 2, Box 3773, 3000 Leuven, Belgium

domain-specific courses requires greater clarity about what set of CT skills could be targeted in domain-specific instruction, how specific subject-matter instruction could be designed considering CT as an integral part of domain-specific instruction, and how best CT outcomes be assessed. Some considerations for the design of CT-supportive learning environments are discussed.

Keywords Domain-general critical thinking · Domain-specific critical thinking · Instructional design · Intervention · Physics

Introduction

Critical thinking (CT) is closely linked with students' in-depth understanding of specific subject-matter content (Williams, Oliver & Stockdale, 2004), improved decision-making with regard to complex real-life problems (Dwyer, Hogan & Stewart, 2012; Halpern, 1993), and more generally with a tendency to become a more active and informed citizen (Halpern, 2014; Tsui, 1999). Various stakeholders in education, such as policy makers, educators, and employers have regarded the development of CT as an essential outcome of undergraduate education (Association of American Colleges and Universities, 2005; Lin, 2014; National Research Council, 1996; Pascarella & Terenzini, 2005). However, efforts to stimulate the development of CT have long been intertwined with controversies over several issues, such as the domain-specificity vs. domain-generality of CT skills (Ennis, 1989; McPeck, 1990b; Smith, 2002), the teaching of CT skills in stand-alone courses vs. within domain specific courses (Ennis, 1989; McPeck, 1990a; Perkins & Salomon, 1989), and the assessment of CT outcomes (Ennis, 1993; Norris, 1989). Taking into account some of the controversies surrounding the teaching of CT skills in the context of higher education, this paper argues that recent developments in instructional design research may have rich implications for designing effective learning environments for CT.

Domain-Specificity and Domain-Generality of CT Skills

Whether CT skills are general, domain-transcending set of skills that can be productively applied in any domain, or are specific to a particular domain, has been highly contentious. On the one hand, some scholars (e.g. Davies, 2013; Halpern, 1998; Kuhn, 1999) claim the existence of a set of CT skills that are general and applicable across a wide variety of domains such as science, history, literature, psychology, and everyday life, on the ground that CT tasks across domains share significant commonalities. On the other hand, some other scholars (e.g. Barrow, 1991; McPeck, 1990b; Moore, 2011) emphasize that the ability to think critically is largely associated with specific criteria within a domain. McPeck, who notably represents the domain-specificity position, argues against the notion of domain-general CT skills on the basis that CT skills required in one domain are different from those required in another (McPeck, 1990b). Supporting this view, Barrow (1991, p. 13) claims that “there are different kinds of concepts that presuppose different types of reasoning,” and thus, CT in one domain is different from CT in another. The counterargument of generalists to the specifists claim has been that of course content and concepts differ from one domain to another, but there are commonalities among thinking practices across domains (e.g. Ennis, 1989; Halpern, 1998).

Coupled with the lack of clear and well-elaborated theory of the concept *domain* (see Ennis, 1989; McPeck, 1990a), the debate over domain-specificity and domain-generality of CT skills has been longstanding. However, there appears to be a recent shift towards a synthesis of the two views (Bailin, Case, Coombs & Daniels, 1999; Davies, 2013; Smith, 2002). First, it has been understood that although content and issues differ from one domain to the next, the synthesis view assumes that there are some commonalities among CT tasks across domains, and thus a set of common CT skills that are applicable across a wide variety of domains do exist. Second, the ability to think critically is recognized to be highly dependent on domain-specific content knowledge, and thus, the synthesis view assumes that an in-depth content knowledge of a particular domain is required for CT competency.

Teaching Students to Think Critically: Review of the Empirical Evidence

The domain-specificity vs. domain-generality debate over CT skills has strongly influenced the approaches to teach CT with respect to regular domain-specific courses. Following the strong generalist position, several studies emphasized the teaching of CT skills separately from regular subject-matter domains (for reviews, see Abrami et al., 2008; Pascarella & Terenzini, 2005). Ennis (1989) refers to such instructional strategies as a *general* approach. Advocates of the general approach argue that CT skills need to be taught in dedicated courses so that they will not be overshadowed by domain-specific content knowledge (Siegel, 1988). However, as CT competency requires in-depth prior domain-specific content knowledge, the general approach had become less dominant in recent years, and CT instruction has mainly focused on embedding CT skills within specific subject domain instruction (Bailin et al., 1999; Smith, 2002).

The notion of embedding CT skills within domain-specific instruction has aroused considerable controversy among researchers and educators since over the past three decades (e.g. Ennis, 1989; Glaser, 1984; Kuhn, 1999; McPeck, 1990b; Moore, 2011; Perkins & Salomon, 1989; Resnick, 1987; Spektor-Levy, Eylon & Scherz, 2009). Some scholars (e.g. McPeck, 1990b; Moore, 2011) assume that meaningful instruction in every subject domain inherently comprises the development of CT skills, and therefore, proficiency in CT skills can be achieved as students construct knowledge of a subject-matter domain without any explicit emphasis on the teaching of general CT skills during instruction. Ennis (1989) refers to such instructional strategies as an *Immersion* approach. Advocates of this approach (e.g. McPeck, 1990a, 1990b) assume that a well-designed subject-matter instruction is sufficient to promote the development of CT skills and equip students to competently perform CT tasks across domains. However, critics of the Immersion approach (e.g. Beyer, 2008; Davies, 2013; Halpern, 1998, 2014) argue that explicit emphasis on general CT skills within specific subject-matter instruction is essential for effective acquisition of CT skills that are transferrable across domains. Ennis (1989) labels such instructional strategies as an *Infusion* approach. Advocates of the Infusion approach argue that when there is an explicit emphasis on why and how a particular CT skill is used within specific subject-matter instruction (e.g. identifying unstated assumptions, assessing the credibility of sources), students become more conscious of when and how that particular skill can be applied in solving CT tasks across domains (e.g. Abrami et al., 2008; Abrami et al., 2015; Beyer, 2008; Halpern, 1998; Kuhn, 1999).

Several researchers have examined the effectiveness of Immersion and Infusion CT instructional approaches for the development of CT skills. Immersion-based instructional interventions that focused on various instructional strategies such as small-group discussion (e.g. Garside, 1996; Stark, 2012), problem-based learning (e.g. Sendag & Odabasi, 2009), repeated practice in higher-order questioning (e.g. Barnett & Francis, 2012), and concept maps (e.g. Wheeler & Collins, 2003) were examined in promoting the acquisition of CT skills. Infusion-based instructional interventions that focused on teacher modeling (e.g. Solon, 2007), role playing (e.g. Toy & Ok, 2012), and coaching (e.g. Bensley & Spero, 2014) were examined in stimulating the development of CT skills. The findings of most of the afore-mentioned studies have been inconsistent. Some of them found that explicit emphasis on CT skills within subject-matter instruction is an effective approach to promote the development of CT skills compared to regular instruction, as measured by domain-general CT tests (e.g. Bensley & Spero, 2014; Dwyer et al., 2012; Solon, 2007), whereas several others reported a non-significant effect (e.g. Anderson, Howe, Soden, Halliday & Low, 2001; McLean & Miller, 2010; Toy & Ok, 2012). Such variability in research outcomes has made it difficult to gain a deeper understanding of the features of Immersion- and Infusion-based interventions for CT. A recent systematic review criticized existing Immersion- and Infusion-based CT intervention studies on the ground that the processes involved in the design and development of the instructional interventions in itself was not sufficiently specified (Tiruneh, Verburgh & Elen, 2014). It is argued in the systematic review that (1) there is little explicit description of the design of previously implemented Immersion- and Infusion-based instructional interventions, and (2) even those explicitly described interventions did not systematically build on the principles of instructional design research. Little consensus exists as a result on the key features of Immersion- and Infusion-based learning environments that are effective for the acquisition of CT skills.

The Assessment of CT Outcomes

Alongside the diverse conceptualization of CT and the longstanding debate on how to teach CT skills, one of the main challenges in CT instruction has been the assessment of CT outcomes. CT has largely been associated with everyday reasoning, and assessment of the effectiveness of Immersion- and Infusion-based instructional interventions has mainly focused on content from everyday life, without reference to domain-specific content knowledge. Researchers have employed various kinds of standardized CT tests that use a broad range of formats, scope, and psychometric characteristics to measure CT outcomes (for reviews, see Ennis, 1993; Halpern, 2015; McMillan, 1987). Most of the tests use content from a variety of everyday life situations with which test takers are assumed to already be familiar, and the tests are labeled as domain-general CT tests (see Ennis, 1993).

Despite the recent shift towards the synthesis of the domain-specificity and domain-general views of CT, the assessment of CT outcomes has thus far mainly focused on domain-general CT skills. The expectation of embedding CT skills within specific subject-matter instruction has been that it will facilitate the acquisition of CT skills that are applicable to a variety of CT tasks within the specific subject-matter domain in question and to CT tasks beyond school subjects (e.g. everyday life situations). Successful teaching of CT skills in coherence with the teaching of domain-specific

content knowledge is in other words expected to result in the development of both domain-specific and domain-general CT skills that are necessary to perform CT tasks requiring a considerable mental activity such as predicting, analyzing, synthesizing, evaluating, and reasoning. However, the experience of evaluating Immersion- and Infusion-based CT interventions for the acquisition of domain-specific CT skills has not been well valued. A few researchers developed and validated CT tests based on content from specific subject-matter domains: the Psychological CT Assessment in the domain of psychology (Lawson, 1999), the Biological CT exam (McMurray, 1991) and the Critical Thinking in Electricity and Magnetism test in the domain of physics (CTEM; Tiruneh, De Cock, Weldeslassie, Elen, & Janssen, 2017). The empirical evidence on whether performance in a domain-specific CT test relates to performance in one of the abovementioned domain-general CT tests has been scant.

The Aim of the Study and Hypotheses

It is argued in this paper that the design of learning environments to embed CT skills within specific subject-matter instruction does not systematically build on instructional design research. Despite the enormous evidence from instructional design research on useful principles to optimize learning and instructional processes, Immersion and Infusion CT instructional approaches particularly have remained underspecified in the CT literature and hence do not sufficiently explain CT research findings. The aim of this study was therefore to examine the effectiveness of systematically designed Immersion- and Infusion-based instructional interventions in promoting the development of domain-specific CT, domain-general CT, and course achievement. In line with recent developments in cognitive psychology (e.g. Merrill, 2002, 2013; van Merriënboer, 1997), learning environments for CT were systematically designed based on empirically valid instructional principles.

For the purpose of this study, CT is viewed from domain-specific and domain-general perspectives. Following the domain-specificity view, we assume that a particular CT task requires domain-specific content knowledge to be competently performed. For example, in view of Halpern's (2014) conceptualization of CT, the use of CT skills makes desirable outcomes more likely, and thus increasing the probability of a desirable outcome requires domain-specific content knowledge. CT skills applied to solve CT tasks that require domain-specific content knowledge are referred to as *domain-specific CT skills*. Unlike specifists, our use of the phrase "domain-specific CT skill" does not suggest that a CT skill employed to competently solve a CT task within a domain applies to that specific domain only. Rather, we are referring to the fact that a CT task may require domain-specific prior content knowledge for it to be competently performed.

In addition, following the domain-general view, we assume that CT tasks across domains share significant commonalities and therefore CT skills can transfer from one domain to another. CT skills that transcend the domain in which they were initially introduced and make desirable outcomes more likely in everyday life are referred to as *domain-general CT skills*.

The design of the Immersion- and Infusion-based instructional interventions in this study focused on a freshman introductory physics course, namely Electricity and Magnetism (E&M), and the following general research question was addressed: What

are the effects of E&M instructional interventions designed based on Immersion, Infusion, and regular instructional approaches on the acquisition of domain-specific and domain-general CT skills and course achievement? In line with the theoretical literature, it was hypothesized that the Immersion and Infusion instructional conditions would result in a significantly higher performance on domain-specific CT, domain-general CT, and course achievement than the regular E&M instruction (Hypothesis 1). The Immersion and Infusion instructional conditions, however, were expected to demonstrate non-significant differences on domain-specific CT and course achievement because both Immersion and Infusion instructional approaches equally target deeper understanding of course content (Hypothesis 2). However, because CT skills were explicitly emphasized in the Infusion condition only, it was hypothesized in line with previous research (e.g. Abrami et al., 2008; Niu, Behar-Horenstein & Garvan, 2013) that the Infusion condition would produce a significantly higher improvement on domain-general CT proficiency than the Immersion condition (Hypothesis 3).

Method

Participants

This study employed a quasi-experimental design involving 147 first-year students with majors in physics, chemistry, or geology. The participants were enrolled in an introductory E&M course at two public universities in Ethiopia. The physics majors at university 1 were purposely assigned into an Infusion group (Infusion-physics, $n = 33$) and the physics majors at university 2 into a control group ($n = 42$). Both the chemistry ($n = 30$) and geology ($n = 42$) majors were at university 1, and each of them was randomly split into two equal groups. Half of the chemistry and geology majors were combined and constituted one group (chem-geo-1, $n = 36$), and the remaining half of each major formed another group (chem-geo-2, $n = 36$). These two groups were randomly assigned to an Infusion (Infusion-chem-geo, $n = 36$) and Immersion conditions (Immersion-chem-geo, $n = 36$). Three participants from the control group and one from the Immersion had to be excluded because of missing posttest data, leaving a final total sample of 143 students. See Table 1 for the distribution of study participants across age and sex.

Description of the Selected Instructional Design Model

In order to design the Immersion- and Infusion-based interventions, we focused on the five CT elements as identified by Halpern (2014): reasoning, hypothesis testing, argument analysis, likelihood and uncertainty analysis, and decision-making and problem-solving. We focused on these CT elements because they are based on recent conceptualization of CT in higher education and comprehensive enough to evaluate students' CT competency (Halpern, 2014).

In line with Halpern's (2014) five elements of CT, we initially identified a list of desired domain-specific and domain-general CT outcomes that our study participants were expected to achieve at the end of the interventions (see Table 2). The First Principles of Instruction model (Merrill, 2002, 2013) was used as a framework to design the Immersion- and Infusion-based E&M interventions because of its

Table 1 Distribution of the participants of the study across age and sex

Group	Number	Age (years)		Sex	
		<i>M</i>	Range	Male	Female
Infusion-physics	33	20.24	18–23	31	2
Infusion-chem-geo	36	20.08	18–22	33	3
Immersion-chem-geo	36	20.09	18–23	33	3
Control	42	20.36	18–22	38	4

comprehensiveness and strong theoretical foundation. This model synthesizes five empirically validated instructional design principles which emerged from research on subject-matter teaching, and offers concrete guidelines to design learning environments for the acquisition of higher-order learning outcomes (Merrill, 2013): Problem-centered, activation, demonstration, application, and integration. The model suggests the use of meaningful and contextually relevant learning tasks and aims to provide students with a variety of learning activities that facilitate the active and constructive acquisition of knowledge and skills. Merrill emphasizes that subject-matter instruction designed on the basis of those instructional principles can result in effective, efficient, and engaging learning, which leads to student acquisition of the knowledge and skills necessary to perform complex tasks (Merrill, 2013).

Among the total 10 chapters included in the E&M course, the intervention focused on the first five chapters: electric field, electric flux, electric potential energy, capacitor and capacitance, and direct current circuits. The content and number of hours allocated for the course were the same for all the Immersion, Infusion, and control conditions.

Description of the Immersion and Infusion Learning Environments

The following is a brief description of the various instructional activities developed and implemented based on the First Principles of Instruction model for both the Immersion and Infusion conditions. Both approaches equally focused on helping students develop deep understanding of E&M content. The Immersion-based E&M instruction engaged students in various domain-specific instructional activities that could result in the achievement of desired domain-specific and domain-general CT outcomes, but without any explicit teaching of general CT skills. In the case of the Infusion-based E&M instruction, however, an explicit emphasis on the desired CT skills was included as an additional layer to the Immersion-based intervention. Each of the five chapters were considered as an integrated whole for the Immersion- and Infusion-based E&M instructions, and lessons progressed from providing to the students relatively simple but meaningful comprehensive E&M tasks to more complex ones. An interdisciplinary team of researchers and two regular E&M teachers collaborated in designing the Immersion and Infusion E&M interventions.

Problem-Centered and Activation Principle Few days prior to the first lesson of each chapter, students were given a meaningful and comprehensive E&M task so that they could conduct an independent inquiry and come up with a brief report that answers the

Table 2 Description of desired domain-specific and domain-general CT outcomes

CT skill	Domain-specific CT outcome	Domain-general CT outcome
Reasoning	In the context of E&M, the student will be able to: <ul style="list-style-type: none"> - Recognize ambiguity of terms - Recognize errors of measurement - Interpret the results of an experiment 	In the context of everyday situations, the student will be able to: <ul style="list-style-type: none"> - Recognize ambiguity of terms - Evaluate/analyze ideas from different perspectives
Thinking as hypothesis testing	<ul style="list-style-type: none"> - Identify important relationships - Examine the adequacy of observations/samples/repetitions of an experiment to draw a conclusion - Check for adequate sample size and possible bias in sampling when making a generalization 	<ul style="list-style-type: none"> - Identify cause and effect relationships - Recognize the need for more information in order to make valid conclusions - Examine the adequacy of observations/samples/repetitions before drawing a conclusion
Argument analysis	<ul style="list-style-type: none"> - Identify key parts of an argument on issues related to E&M - Infer a correct statement from a given data set - Criticize the validity of generalizations drawn from the results of an experiment 	<ul style="list-style-type: none"> - Identify key parts of an argument: e.g. given a conclusion, identify the reason(s) that support the conclusion - Infer a correct statement from a given data set - Criticize the validity of generalizations
Likelihood and uncertainty analysis	<ul style="list-style-type: none"> - Predict the probability of events (but understand the limits of extrapolation) - Identify assumptions (e.g. recognize what assumptions have to be maintained in generalizations drawn from experimental results) - Understand the need for additional information in making decisions - Make valid predictions 	<ul style="list-style-type: none"> - Understand the probability and likelihood of an event occurrence - Identify assumptions - Make valid predictions: what-if questions - Understand the need for adequate sample sizes
Problem-solving and decision- making	<ul style="list-style-type: none"> - Identify the best among a number of alternatives in solving E&M-related problems - Examine relevance of procedures in solving scientific problems - Evaluate solutions to an E&M-related problem - Make sound, evidence-based decisions - Use analogies to solve E&M-related problems 	<ul style="list-style-type: none"> - Identify the best option from a number of alternatives in solving everyday problems - Decide on the validity of a particular scientific explanation when applied to new situations - Examine the relevance of the procedures in solving problems - Use analogies to solve problems - Develop reasonable, creative solutions to a problem

comprehensive task. Preparing the report was a requirement to all the students, but it was not graded. In the first lesson of each chapter, students were asked to discuss in small groups on their reports, and afterwards, the teacher modeled an epitomic version of the answer to the comprehensive task. The main subtopics within the chapter were subsequently introduced and the teacher asked a few oral questions that could activate students' prior knowledge on the topics. What was different in the Infusion E&M lesson is that the teacher explicitly introduced in the initial lesson of the course that students would be guided to learn some useful CT skills as part of the course. In addition, the Infusion teacher made explicit reference to one of the desired CT skills at the beginning of each of the five chapters while modeling solutions to comprehensive tasks.

Demonstration Principle Each new topic was initially explained and adequate information was presented during lessons in both the Immersion and the Infusion lessons. The teachers then modeled by thinking aloud how the earlier presented information could be used in solving the E&M tasks. What was different in the Infusion environment is that the teacher made an explicit reference to a particular thinking strategy while modeling the solution to an E&M problem by asking questions such as “do I have now sufficient information to make a sound conclusion?” and “how do I relate this strategy with the CT skills I introduced at the beginning of the chapter?”.

Application Principle Students in both the Immersion and Infusion lessons were asked to solve numerous E&M problems that required them to interact with one another both in solving and evaluating solutions, and explaining their solutions to group members. The teachers in both conditions coached the problem-solving activities of the students, provided corrective feedback when required, and facilitated small group discussions. In the Infusion group, however, the teacher kept students focused throughout on how a particular CT skill could be applied to solve the E&M problems, and how that CT skill could be applied in different settings. The Infusion teacher acted as a group member during small group discussions and asked some probing questions such as “how did you apply the principles of inductive reasoning in solving this problem?” and “how could this strategy be used to solve problems in other courses?”.

Integration Principle Both the Immersion and Infusion E&M instructions focused on encouraging students to reflect in small groups on their E&M problem solutions, and occasionally, students were encouraged to present their solutions to the whole class. At the end of each chapter, students were also required to refer back to the comprehensive task given at the beginning of the chapter and to give a detailed and complete solution to the task. Moreover, students were induced to prepare a brief summary of the important E&M concepts learned within a chapter by using concept maps. In the Infusion learning environment, however, students in addition were encouraged to prepare summary of the learned CT skills within each chapter.

Description of the Regular E&M Instruction (Control Group)

The E&M instruction for the control condition was designed and developed by the regular teacher at university 2. Teaching method in Ethiopian higher education is

mainly traditional, which is characterized as less engaging, highly dominated by the teacher, limited collaboration among students, and little practice in answering higher-order thinking questions (Asgedom et al., 2009). Most of the instructional time in Ethiopian higher education involves the teacher lecturing to students, and assignments are largely end-of-the-chapter type homework problems with short quantitative answers.

As noted above, the content and lesson durations for the control condition were the same as that for the Immersion and Infusion conditions. Efforts were made also to carefully control students' time on task as far as the E&M course was concerned. It should be noted that the Immersion and Infusion groups were required to solve comprehensive E&M problems ahead of the first lesson in each chapter and submit brief reports. To counterbalance the time on task, students in the control group were in return given reading assignments of selected topics a few days prior to the beginning of each chapter and they were required to submit summary reports during the first lesson of each chapter. To obtain an overview of the instructional processes, the first author observed two of the control group's lessons, and interviews were conducted with the E&M teacher on three separate occasions: at the beginning of the semester, a month after the semester started, and at the end of the intervention. A detailed analysis of the classroom observations and interview data revealed the precise differences of the regular E&M instruction with respect to the First Principles of Instruction model (see Table 3 for a detailed description of the differences between the Immersion, Infusion, and control learning environments).

Instruments

The HCTA (Halpern, 2015) The Halpern Critical Thinking Assessment (HCTA) was administered both as a pretest and a posttest to measure the acquisition of domain-general CT skills. The test focuses on the five elements of CT that are targeted in the intervention, and consists of 20 items based on a variety of real-life problems such as health, education, politics, and social policy. For each CT element, four items were included, and each item is followed by questions that require respondents to first provide brief constructed responses (constructed-response items) and to subsequently select answers from a short list of alternatives (forced-choice items). The internal consistencies for both formats of the HCTA in the present study were acceptable based on the guidelines by Nunnally (1978): Cronbach's $\alpha = 0.72$ for the pretest HCTA constructed-response and .71 for the pretest HCTA forced-choice formats, $N = 147$; .74 for the posttest HCTA constructed-response and .72 for the posttest HCTA forced-choice formats, $N = 143$. Both formats were scored based on the scoring guidelines provided by Halpern (2015). The maximum score expected for both the forced-choice and constructed-response formats is 154.

CTEM (Tirunch et al., 2017) The CTEM was administered to measure domain-specific CT proficiency in line with the desired domain-specific CT outcomes described in Table 2. The test consists of 20 items: two of which are forced-choice and the remaining are constructed-response format items. The test authors designed the CTEM items to mirror the five CT elements identified in the HCTA, but focus on E&M content (see Fig. 1 for sample HCTA and CTEM items). The CTEM was administered in the

Table 3 Description of the Immersion, Infusion, and regular E&M instructions in relation to the First Principles of Instruction model

The design principle	Immersion-based E&M instruction	Add-ons to the Infusion E&M instruction	Regular E&M instruction
<p><i>Problem-centered principle:</i> Learning is promoted when learners acquire knowledge in the context of real-world problems. Problems need to be comprehensive, challenging and representative of the problems learners will encounter in real-life (Merrill, 2013).</p>	<p>For each chapter, relatively complex, meaningful, and comprehensive problems were carefully designed by seeing each chapter as a mini-course (based on the suggestion by Merrill). An attempt was made to keep the tasks relevant to the lives of students and thus make them more motivating.</p>	<p>For about 20 min, students were introduced explicitly at the beginning of the course regarding the desired CT outcomes. The orientation mainly focused on explaining what it means to think critically and a brief introduction of the five targeted CT skills.</p>	<p>Instruction was primarily “topic-centered.” At the beginning of a new chapter, the teacher presented information related to that chapter (or subtopic). Students were sometimes shown solutions to one or two textbook problems related to the newly presented information. At the end of the lesson, students were given selected textbook problems as homework assignments. The E&M problems were not designed to echo real-world problems.</p>
<p><i>Activation principle:</i> Learning is promoted when learners activate existing knowledge as a foundation for new knowledge and skill. Instruction should not start with abstract representations for which learners have insufficient background (Merrill, 2013).</p>	<p>Various activities that helped learners make meaningful connections between newly acquired and their prior knowledge were carefully prepared in advance and implemented during instruction. For example, learners received questions about a specific topic that aimed to relate the concepts of the new topic to their prior knowledge, and they were required to share their answers with other learners (peer sharing).</p>	<p>The instructional activities involving the activation principle were the same in the Infusion condition.</p>	<p>There were no systematic attempts to activate learners’ prior knowledge before information on a new topic was presented. When a lesson on a new topic began, the instructor usually started by briefly explaining the topic and subsequently presenting detailed information on the topic. Sometimes, the teacher encouraged students to tell what they remembered of the previous lesson, but no further prompts were offered to help students describe the preceding lessons in detail.</p>
<p><i>Demonstration principle:</i> Learning is promoted when learners observe a demonstration of the knowledge to be learned that is consistent with the type of content being presented (Merrill, 2013).</p>	<p>We mainly changed the textbook’s standard numerical E&M problems into more qualitative/conceptual problems. An attempt was made to qualify the tasks so that the desired CT skills could implicitly</p>	<p>Explicit reference was made by the teacher on the targeted CT skills when modeling the solutions to various E&M problems (e.g. do I need to draw a free body diagram? what can I conclude? what can</p>	<p>A great deal of information was presented, but was on telling rather than both telling and showing the information. After he introduced a lesson, the teacher presented detailed information on the topic, but he</p>

Table 3 (continued)

The design principle	Immersion-based E&M instruction	Add-ons to the Infusion E&M instruction	Regular E&M instruction
	<p>be modeled by the instructor when presenting the solutions (e.g. what can we conclude from the answer? what other options are there to solve this problem?; what other information do we need to solve this problem?).</p> <p>Relevant and challenging E&M tasks were designed that created several opportunities for the students to engage in applying newly presented information. When students were engaged in solving problems, activities that facilitated teacher coaching and guidance were clearly described and implemented. For instance, the teacher provided partial solutions, halted at each group and observed students' discussions, provided hints as needed, acted as group members and asked thought-provoking questions, encouraged students to formulate questions using specific verbal prompts, and facilitated discussion among group members.</p>	<p>I say about the relationship between these two variables? What kind of reasoning am I making? Inductive or deductive reasoning? This is an example of inductive reasoning; this is an example of argument analysis).</p> <p>The instructional activities were largely similar to the Immersion group, but the teacher kept the students focused on how a particular CT skill can be applied to solve the E&M problems.</p>	<p>did not adequately show how the presented information might be used to solve a new problem. Tasks that might have facilitated demonstration of the newly presented information were not systematically designed in advance.</p> <p>Students were not engaged in applying the newly presented information to solve new and meaningful E&M problems; rather, the instructor gave them traditional end-of-chapter problems as homework assignments. Moreover, there was no dedicated time for students to practice solving as many practical E&M problems as possible during the lessons. Even when they were asked questions, the questions focused on recalling information and did not invite further elaboration and explanations from the students.</p>
<p><i>Application principle:</i></p> <p>Learning is promoted when learners engage in application of their newly acquired knowledge that is consistent with the type of content being taught (Merrill, 2013).</p>	<p>Activities that encourage students to present their solutions either to group members or the whole class were designed, and both peer and instructor feedback were offered. Representatives from groups were sometimes asked to present in front of the full class. Students in other groups</p>	<p>The instructional activities were similar to that of the Immersion condition, but students in this condition were required to prepare a summary of the learned CT skills and how those skills were applied in solving the E&M problems.</p>	<p>Students usually did not have the opportunity to present and defend their solutions to group members or the full class. Interaction between the students during the lessons was very limited: they did not engage in exchanging ideas and explaining solutions to problems between themselves or the instructor.</p>
<p><i>Integration principle:</i></p> <p>Learning is promoted when learners integrate their new knowledge into their everyday lives by being required to reflect on, discuss, or defend their new knowledge via peer collaboration and peer critique (Merrill, 2013).</p>			

Table 3 (continued)

The design principle	Immersion-based E&M instruction	Add-ons to the Infusion E&M instruction	Regular E&M instruction
	were encouraged to ask questions, and the student presenters were asked to defend their solutions when challenged by their classmates or the instructors.		Most importantly, the E&M problems did not usually invite students to apply what was learned in solving new and meaningful E&M problems.

Sample HCTA item

Four patients were waiting to see a doctor who specializes in treating headaches. Three of the four patients were women, which led the male patient to declare that more women seek medical help for their headaches than men.

A. Is this a reasonable conclusion based on the people waiting to see this doctor?

☐ Yes

☐ No

B. Please explain your answer

Sample CTEM item

The electrical resistivity of different materials is measured as a function of temperature. The results are given in the table below.

Temperature (K)	Resistivity Aluminum ($10^{-8} \Omega\text{m}$)	Resistivity Gold ($10^{-8} \Omega\text{m}$)	Resistivity Iron ($10^{-8} \Omega\text{m}$)	Resistivity Copper ($10^{-8} \Omega\text{m}$)
1	0.0001	0.0220	0.0225	0.002
10	0.000193	0.0226	0.0238	0.00202
100	0.442	0.65	1.2800	0.348
200	1.587	1.462	5.2000	1.046
300	2.733	2.271	9.9800	1.725
400	3.87	3.107	13.1000	2.402
500	4.99	3.97	23.7000	3.09
600	6.13	4.87	32.9000	3.792
700	7.35	5.82	44.0000	4.514
800	8.7	6.81	57.1000	5.262

Based on these measurements, can you conclude that ‘resistivity increases with increasing temperature’?

☐ Yes

☐ No

Explain your answer.

Fig. 1 Sample HCTA and corresponding CTEM items

present study as a posttest only. Because the test requires prior knowledge of E&M, we felt that it was reasonable to administer the test only at the end of the intervention. In return, however, the grade 12 university entrance national exam scores for physics were used to control for physics prior knowledge of the study participants. The internal consistency of the CTEM (Cronbach’s $\alpha = 0.73$, $N = 143$) for the present study was found to be acceptable (Nunnally, 1978). The CTEM test scoring guide, prepared in line with the HCTA scoring guide, was used to score the CTEM items. The maximum score expected for the CTEM test was 63. See Fig. 2 for sample student responses to a CTEM item and corresponding awarded scores.

Course Achievement Test A teacher-made test, similar to end-of-course exams, was administered to all the study participants at the end of the interventions. The achievement test consists of 22 items: 19 of which are forced-choice, and the remaining are constructed-response format items. The maximum score expected in this test was 36. Representative items from all the five chapters focused on in the interventions were developed by a physics teacher from university 1. The test developer had taught the course for several years, but was not involved in the present study. One of the co-authors, a physics professor, reviewed the suitability of the test to measure the desired learning outcomes of the five chapters, and modifications were made in collaboration with the test developer. All the Infusion, Immersion, and control group teachers who participated in the present study were also asked immediately after the end of the intervention to evaluate the suitability and clarity of the items. Minor adaptations were made based on the feedback of the participating teachers. The internal consistency of this test (Cronbach's α) was 0.67, $N = 143$. It should be noted that the course achievement test was teacher-made and did not pass through rigorous validation procedures.

Procedure

To compute the interrater agreements of the constructed-response format items of all the three tests, 40 randomly selected test papers (10 from each condition) were scored independently by two different raters using the same scoring guides. Paired sample t test was computed to examine the effect of the rater on the mean scores of each of the constructed-response format items of the HCTA, the CTEM, and the course achievement. The results indicated no statistically significant differences between the scores allocated by the two raters to each item of the three tests ($p > .05$).

The study participants completed the paper version of the HCTA at the first day of the intervention as a pretest and a week after the end of the intervention as a posttest. The regular procedures for administering the HCTA were followed: participants were asked to first answer the HCTA constructed-response format items and subsequently the HCTA forced-choice format items. The CTEM test was however administered as a posttest only. Participants in all the conditions completed the posttests in multiple sessions spread over separate days: the CTEM was administered firstly, followed by the course achievement test, and finally the HCTA. All tests were administered in a classroom setting. The CTEM and course achievement tests each lasted between 50 and 75 min, and the HCTA (both formats) between 70 and 90 min.

Implementation of the Experimental Interventions

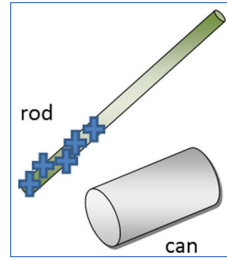
The designed interventions for all the conditions were implemented in the 2014–2015 academic year over 8 weeks with three lessons of 2 h each per week. The Immersion and Infusion teachers had collaborated during the design and development phases of the interventions. In order to control for the teacher effect, teachers who had the same education level and equivalent years of teaching experience were involved in implementing the interventions. The Immersion and Infusion teachers received all the necessary information regarding the purpose of the interventions and what they were

Sample CTEM item:

Hanna conducts the following experiment: she brings a positively charged rod close to a metal can. Doing the experiment shows that the can is attracted to the rod. Hanna is puzzled with the result of her experiment. She expected the negative electrons on the metal would be attracted to the rod while the positive nuclei are repelled, and opposite forces cancel out, which would mean that the can remains at rest.

How can you make Hanna's argument consistent with the experiment?

Indicate all the possible explanations.



Sample student responses for the item and the scores awarded (Item weight = 4 points):

Ideal complete answer expected from a student:

The positively charged rod draws the loosely bound electrons and accumulates them at the side of the can closest to the rod while leaving the other side positively charged. Because the distance between the rod and the negatively charged side of the can is smaller than the distance between the rod and the positively charged side of the can, the attractive force between the rod and the can is larger than the repulsive force between them. According to Coulomb's law, the force is inversely proportional to the square of the distance. Therefore, the net force on the can is attractive.

Student 1:

The amount of positive kernels and negative electrons are not equal, the rod has much more positive kernels thus the net force is not zero.

➔ Awarded 0 points

Student 2:

The electrons move through the 2 bodies. If we approach the can with the rod, more electrons will move to the surface causing the rod and the can to attract each other. We get a redistribution of charges. As the rod is positively charged, there are more protons in the rod than electrons.

➔ Awarded 1 point (motion of electrons = 1)

Student 3:

The electrons in the can move towards the rod, so the average distance between the rod and the electrons is smaller than the average distance between the rod and positive particles. There is a non-zero resultant force.

➔ Awarded 2 points (motion of electrons = 1, distance mentioned = 1)

Student 4:

Because negative electrons are attracted and positive ions are repelled, the can will have a positive and a negative side. As the electric force decreases with distance ($F \sim 1/r^2$), the negative side will be attracted more strongly than the positive side is repelled. The can will move towards the rod because there is net force towards the rod.

➔ Awarded 4 points (motion of electrons = 1, $F_{\text{attract}} > F_{\text{repel}} = 1$, $F_{\text{Coulomb}} = 2$)

Fig. 2 Sample student responses for a CTEM item

required to do in implementing the lessons as designed, and the first author monitored the execution of the interventions. Two major elements were emphasized while monitoring the fidelity of implementation of the interventions: (a) the extent to which the implementation corresponds to the design and (b) participants' responsiveness to the

newly designed instructional activities. Regarding the first element, the observation disclosed that the interventions were basically implemented as designed. Some factors can be mentioned that optimized the implementation as per the design. First, the experimental teachers participated actively during the design phase of the interventions, and also received training right before the implementation aimed at giving an overview of the features of the interventions. Second, a lot of effort was made during the design and development phases to describe in detail the various components of the learning environment (student and teacher activities, tasks, prompting questions, etc.). Third, the first author provided feedback (during postlesson discussions) that maximized the implementation of the interventions as designed throughout the experiment.

Regarding the second element (study participants' responsiveness), our observation and interview data revealed some useful information. The experimental teachers acknowledged the newly designed learning environments encouraged the participants to be engaged actively during the E&M instruction. However, most of the instructional activities were new and the participating students appeared to be confused particularly during the first 2 weeks of the intervention. It was observed that the experimental teachers provided the necessary guidance and the students started to comply with the instructional activities after the second week of the intervention.

Analyses

Despite the absence of complete random assignment of participants to the different conditions, the groups were comparable in a number of important features. First, they were all freshmen and there were no marked differences on average age. They had also similar educational backgrounds and no significant differences in prior physics knowledge. A one-way analysis of variance (ANOVA) revealed no significant difference between the four groups in their physics prior knowledge (as measured by the national college entrance exam for physics), $F(3, 139) = .064, p = .97$, and pretest HCTA scores, $F(3|139) = .191, p = .90$. The two Infusion groups, namely *Infusion-physics* and *Infusion-chem-geo*, had participated in exactly the same E&M instructional interventions and were taught by the same teacher. Because initial comparisons of prior physics knowledge and pretest HCTA proficiency revealed no significant differences between the four groups, we merged the *Infusion-physics* and *Infusion-chem-geo* groups into one *Infusion* group on the postintervention comparisons. The research hypotheses were tested by using type III sums of squares, which weighs the sample means equally irrespective of differences in sample sizes (Tabachnick & Fidell, 2007).

The three main outcome variables were the CTEM, the HCTA, and the course achievement scores. The data for these variables were initially screened for missing values, outliers, and normality of distributions separately within each group. The proportions of missing values per item for all the variables were very limited (<5%) and randomly scattered over each of the outcome variables. Mean substitution was employed to impute missing values. Tests of assumptions for normality for the CTEM, HCTA, and course achievement scores were done using visual inspection of the boxplots, histograms, and Q-Q plots. The boxplots for all the variables suggested a relatively normal distributional shape (with no outliers) of the residuals. The histograms and Q-Q plots also suggested that normality was reasonable.

Results

The Effect of the Instructional Interventions on the Outcome Variables

A one-way multivariate analysis of variance (MANOVA) was first performed on the means of the three outcome variables. However, prior to conducting the MANOVA, a series of Pearson correlations were performed between all the three outcome variables in order to test one of the MANOVA assumptions that the outcome variables would be moderately correlated with each other (Tabachnick & Fidell, 2007). As can be seen in Table 4, all the outcome variables were moderately correlated with each other, suggesting the appropriateness of a MANOVA. In addition, Box's test was computed to check the assumption of equality of covariance matrices. The result showed that the covariance matrices between the groups were assumed to be equal ($p = .14$) for the purposes of the MANOVA.

A one-way MANOVA was conducted to test the first hypothesis that there would be one or more significant mean differences between the Immersion, Infusion, and regular E&M instructional conditions on domain-specific CT proficiency, domain-general CT proficiency, and course achievement scores. Using the Wilks' statistic, there was a significant effect of the instructional conditions on the three outcome variables, $\Lambda = 0.74$, $F(6, 276) = 7.31$, $p < .001$. The multivariate effect size was estimated at 0.137, which implies that 13.7% of the variance on combination of the outcome variables was accounted for by the instructional interventions. The homogeneity of variance assumption was separately tested for all the outcome variables prior to conducting a series of follow-up ANOVAs. Based on a series of Levene's F tests, the homogeneity of variance assumption was considered satisfied for the CTEM ($F(2, 140) = 0.452$, $p = .64$), posttest HCTA ($F(2, 140) = 0.31$, $p = .74$), and course achievement ($F(2, 140) = 0.63$, $p = .54$). The one-way ANOVAs on the outcome variables revealed significant intervention effects only on domain-specific CT proficiency, $F(2, 140) = 13.54$, $p < .001$, $\eta_p^2 = 0.162$, and course achievement, $F(2, 140) = 12.48$, $p < .001$, $\eta_p^2 = 0.151$, but not on domain-general CT proficiency, $F(2, 140) = .241$, $p = .79$. The effect sizes associated with the statistically significant effects are considered large based on Cohen's (1988) guidelines, with the instructional interventions accounting for 16.2% of the variance on domain-specific CT and 15.1% on course achievement.

In order to examine whether there was significant pretest-posttest improvement on domain-general CT outcomes across the three instructional conditions, a mixed-design ANOVA was conducted. The results revealed non-significant interaction between the

Table 4 Pearson correlation coefficients, means, and standard deviations of outcome variables

Variable	1	2	3	4	<i>M</i>	<i>SD</i>
1. Course achievement	—				23.29	3.19
2. CTEM	0.38*	—			31.35	5.67
3. Pretest HCTA	0.05	0.12	—		79.99	7.85
4. Posttest HCTA	0.28*	0.41*	0.38*	—	82.38	7.12

$N = 143$

* $p < .001$

testing period (pretest-posttest) and the instructional conditions (Immersion-Infusion-control), $F(2, 140) = .162, p = .85$. This implies that the domain-general CT scores for either the Immersion or the Infusion condition did not show significant pretest-posttest improvements compared to the control condition. The descriptive statistics associated with all the variables across the three instructional groups are reported in Table 5.

Comparison of the Groups on Domain-Specific CT Skills and Course Achievement

In order to examine the pairwise differences across the means of the three instructional conditions on domain-specific CT and course achievement, the ANOVA was followed up with the Hochberg's GT2 post hoc test. This test is selected as the sample sizes were different across groups (Field, 2009). For domain-specific CT proficiency, the results revealed statistically significant differences between the Infusion and control groups, $p < .001, d = 1.07$, and the Immersion and the control groups, $p = .015, d = .69$. However, the test indicated that the domain-specific CT proficiency scores did not differ significantly between the Immersion and Infusion groups ($p = .196$). The effect sizes associated with the statistically significant differences are considered moderate to large based on Cohen's (1988) guidelines. For course achievement, the results revealed statistically significant differences between the Infusion and the control groups, $p < .001, d = .96$, and the Immersion and control groups, $p = .001, d = .86$. However, the difference between the Infusion and Immersion groups was not statistically significant ($p = .97$). The effect sizes associated with the statistically significant differences are considered large based on Cohen's (1988) guidelines.

Discussion

It was argued in this study that instructional interventions for CT need to be systematically designed based on empirically valid instructional principles and that the evaluation of the effects of those interventions needs to focus on both domain-specific and domain-general CT skills. Accordingly, Immersion- and Infusion-based E&M instructional interventions were designed based on the First Principles of Instruction model and evaluated with respect to the acquisition of domain-specific CT skills, domain-general CT skills, and course achievement.

Table 5 Descriptive statistics for the CTEM, HCTA, and course achievement scores across the instructional conditions

Group	CTEM		HCTA				Course achievement	
			Pretest		Posttest			
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Immersion ($n = 35$)	31.34	5.31	80.43	9.15	82.17	7.69	23.89	2.99
Infusion ($n = 69$)	33.32	5.43	80.06	7.05	82.80	6.81	24.13	2.88
Control ($n = 39$)	27.87	4.76	79.46	8.11	81.85	7.27	21.28	3.07

The findings revealed that both the Immersion and Infusion E&M instructional conditions significantly outperformed the regular E&M instruction condition on domain-specific CT proficiency and course achievement, but not on domain-general CT proficiency. The findings suggest that engaging students with systematically designed instructional activities that give either an implicit or explicit emphasis on desired CT skills can significantly foster the acquisition of domain-specific CT skills and course achievement. These findings partially support Hypothesis 1 and are consistent in general with the CT theoretical literature that argues for the effectiveness of a well-designed subject-matter instruction in enabling students solve domain-specific CT tasks (e.g. Glaser, 1984; Perkins & Salomon, 1989; Resnick, Michaels & O'Connor, 2010; Smith, 2002) and achieve better on course content measures (Beyer, 2008; Resnick, 1987; Williams et al., 2004).

Consistent with Hypothesis 2, post hoc analyses indicated that domain-specific and course achievement scores did not significantly differ between the Immersion and Infusion conditions. As noted above, both the Immersion and Infusion conditions equally focused on students' in-depth understanding of the E&M content (i.e. lessons were carefully designed based on the First Principles of Instruction model for the two conditions), and CT skills were integral components of the E&M instructional activities in both cases. We argued earlier that (domain-specific) CT skills can and should be essentially targeted in well-designed subject domain instruction, and the fact that the Immersion group demonstrated domain-specific CT proficiency equally to that of the Infusion group was consistent with our expectation. Moreover, the lack of significant difference in course achievement between the Immersion and Infusion conditions was interesting because it reveals vital evidence that explicit focus on selected CT skills within the regular course instruction did not function at the cost of students' content knowledge of E&M.

Contrary to our expectation, however, the findings revealed non-significant differences between the Immersion, Infusion, and control conditions on the acquisition of domain-general CT skills. Because an explicit emphasis was given on selected CT skills in the Infusion condition, we expected that the Infusion condition would produce significantly higher improvement on domain-general CT skills compared to the control and Immersion conditions (Hypothesis 3). However, improvements on the acquisition of domain-general CT skills did not significantly differ depending on the instructional conditions they were engaged with. It was indeed unexpected that the Infusion condition failed to result in significantly higher pretest-posttest improvement even compared to the control condition. This finding is contrary to previous findings (e.g. Abrami et al., 2008; Angeli & Valanides, 2009; Bensley & Spero, 2014; Beyer, 2008) that showed explicit emphasis on CT skills within specific subject-matter instruction significantly fosters the development of domain-general CT skills compared to "regular" instruction.

A number of reasons may explain why the Infusion E&M instructional condition did not result in a significantly higher improvement on domain-general CT skills compared to both the Immersion and control instructional conditions. One may be related to the design features of our Infusion-based E&M intervention. The CT skills were probably not sufficiently explicit in the Infusion lessons and were overshadowed by the E&M content. It should also be noted that our intervention targeted only 50% of the E&M content, which was implemented within 8 weeks period. Perhaps the scope and duration of the intervention was highly restricted to produce CT skills that can transfer across domains. Besides, given that the participating students lack prior experience

with the instructional methods introduced in the interventions, it might take some more time before the benefits of the newly designed interventions become noticeable. In addition, the other domain-specific courses in which the study participants were concurrently enrolled during the intervention might not be designed optimally. This may have resulted in limited opportunities for students to extensively practice the desired CT skills in solving various thinking tasks in other domains, and thus restricted the transfer of CT skills across domains. This finding adds to the existing theoretical as well as empirical evidence (e.g. Anderson et al., 2001; Halpern, 2014) that suggests that the acquisition of transferrable CT skills can be achieved mainly through interventions that involve an extended duration and coverage of a large number of courses. Future studies may focus on a more intensive and comprehensive interventions that involve an extended duration and with the inclusion of more than one domain-specific courses.

Another possible explanation for the non-significant effect may relate to the domain-specificity and domain-generality debate over CT skills. As indicated earlier, *specificists* (e.g. McPeck, 1990b) argue against the existence of domain-general CT skills on the basis that CT is always thinking about specific subject domain. We could argue that our finding is consistent with some of the theoretical claims regarding the domain-specificity of CT (e.g. McPeck, 1990b; Moore, 2011). The participants in the Immersion and Infusion groups had demonstrated significant improvement in the targeted domain-specific CT outcomes. However, despite the explicit emphasis on desired CT skills in the Infusion condition, there was no evidence for transfer of the acquired domain-specific CT skills to everyday CT tasks. Advocates of the *specificists'* view may see this lack of transfer as one indication towards the domain-specificity of CT skills. *Generalists* may, however, argue that students' failure to transfer the acquired domain-specific CT skills may have to do with the absence of explicit emphasis on the teaching of CT skills in stand-alone courses.

A third possible explanation may relate to the issue of the assessment of CT outcomes. Following the synthesis of the *generalists'* and *specificists'* views, we assumed that training students systematically to solve various domain-specific CT tasks with an explicit focus on selected CT skills would adequately equip them to solve CT tasks across domains including everyday life. In administering the HCTA, the goal was therefore to examine the extent to which acquired CT skills within domain-specific instruction would transfer to a different domain: everyday reasoning. The HCTA items mainly reflect common experiences across cultures in industrialized societies (Halpern, 2015), and it is possible that our study participants in Ethiopia may have lacked adequate prior knowledge of the content used to prepare the HCTA items. The failure to transfer the acquired domain-specific CT skills to everyday reasoning tasks may have therefore originated from the HCTA itself. In evaluating the effectiveness of CT-supportive instructional interventions on the acquisition of domain-general CT skills, an important issue for future studies would therefore be to make sure that domain-general CT tests actually reflect common everyday problems to study participants.

Study Limitations

Given its quasi-experimental nature, there are some limitations to this study that the reader needs to be aware of in interpreting the findings. First, the Immersion, Infusion, and control conditions were taught by three different teachers, and the control condition

was located in a different university. Attempts were made to minimize the effects of some of the confounding variables with respect to the teacher and the institution. For instance, all the participating teachers had the same education levels and equivalent years of teaching experience, and efforts were also made to closely monitor the implementation of the lessons at all the instructional conditions as per the design. We could argue that the participation of different teachers to each of the instructional conditions and the fact that the control condition was located in a different university were beneficial to examine more accurately the true effects of the instructional interventions. If we had assigned the same teacher to all the three conditions, the learning environments might have been contaminated by the teacher's training and experience in one of the instructional conditions. Besides, the fact that the control group was located in a different university was advantageous as it eliminated possible contact among students assigned in the control and experimental conditions. Second, it should be noted that the interventions were implemented and evaluated in ecologically valid instructional settings. We were convinced that instructional interventions must compete for success in the disarray that constitutes the daily classroom life of the students and the teacher (e.g. Brown, 1992). We faced some challenges as a result in implementing the E&M lessons as designed. We were asking the Immersion and Infusion teachers to adopt instructional approaches that were not very familiar to them. Although the experimental teachers received training and had collaborated during the design phase of the interventions, it was observed that the implementation of the interventions put huge burden both to the teachers and students. Probably the teachers' and students' limited prior acquaintance to such learning environments may have influenced the ideal implementation of the instructional interventions as per the design. Third, the participating students were from three separate domains: physics, chemistry, and geology. Some critics may object the findings of the present study by singling out that we compared three seemingly different groups of students. We acknowledge this as a limitation, but it should be noted that the study participants from the three separate majors were comparable in terms of their prior physics knowledge and pretest domain-general CT scores.

Conclusions

The findings of this study imply considerable practical and theoretical significance for CT and instructional design research. The findings add to the literature base by specifying and elaborating the design of Immersion- and Infusion-based instructional interventions for the acquisition of both domain-specific and domain-general CT skills. Desired domain-specific and domain-general CT outcomes were initially specified, various domain-specific instructional activities based on sound instructional principles were designed, and the effectiveness of the interventions was examined by using various tests. Through this study, we hope to have demonstrated how empirically valid instructional principles can be translated into prescriptions for everyday classroom activities focusing on an Immersion and Infusion CT instructional approaches. The literature largely depicts CT as an elusive concept with little direction on how to translate the diverse views into CT instructional practices. Acknowledging the longstanding controversies involved in defining, teaching, and assessing CT, efforts were made in this study to show how CT can be handled as an integral part of the

domain-specific instruction in which students are being enrolled. Generally, we have demonstrated that embedding CT instruction in domain-specific courses requires greater clarity about what CT is, what set of CT skills could be targeted in domain-specific instruction, how specific subject-matter instruction could systematically be designed considering CT as an integral part of domain-specific instruction, and how best CT outcomes be assessed. Particularly, we have indicated that a systematic approach to embedding CT instruction in domain-specific courses includes formulating and answering the following questions: (a) what does it mean to think critically in a particular domain?, (b) what instructional principles are relevant to achieve desired domain-specific CT outcomes, and how do we translate those instructional principles into usable instructional activities?, and (c) how do we accurately measure for the acquisition of both domain-specific and domain-general CT skills? We assume that answering the aforementioned questions in any effort to embed CT instruction in domain-specific courses may be an effective approach to address the challenges of CT development. The study has clearly demonstrated that students can be successfully guided to acquire and use CT skills at least in the boundaries of their domain. However, the exact features of learning environments to facilitate the transfer of acquired domain-specific CT skills across domains (e.g. to solve domain-general CT tasks that do not require specialized content knowledge) remain unclear. In sum, we hope to have shown in this study that designing instructional interventions systematically and transparently constitutes a promising practice to research on the integration of CT skills within specific subject domains. It is essential that future work continues to explore the effectiveness of such systematic approach to design CT supportive learning environments in promoting the development of both domain-specific and domain-general CT skills.

Acknowledgements We would like to gratefully acknowledge the assistance given to us by the staff and students from the two participating universities. We would like to thank particularly Abraham Kassa, Solomon Ashagre, Bantayehu Aderaw, and Zinaye Tefera for their support during the entire implementation period. A lot of thanks as well to Ataklti Weldelessie for the valuable contribution during the design phase of the various E&M activities, Dawit A. Getahun for his advice with respect to data analysis, and Jan Sermeus for his help with the scoring of the tests.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. *Review of Educational Research*, 85(2), 275–314. doi:10.3102/0034654314551063.
- Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M., Tamim, R., & Zhang, D. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research*, 78(4), 1102–1134. doi:10.3102/0034654308326084.
- Anderson, T., Howe, C., Soden, R., Halliday, J., & Low, J. (2001). Peer interaction and the learning of critical thinking skills in further education students. *Instructional Science*, 29, 1–32.
- Angeli, C., & Valanides, N. (2009). Instructional effects on critical thinking: Performance on ill-defined issues. *Learning and Instruction*, 19(4), 322–334. doi:10.1016/j.learninstruc.2008.06.010.
- Asgedom, A., Shibeshi, A., Mekonnen, D., Nega, M., Semela, T., Leka, W., & Endawoke, Y. (2009). *Quality of higher education in public institutions in Ethiopia*. Addis Ababa: Forum for Social Sciences.

- Association of American Colleges and Universities. (2005). *Liberal education outcomes: A preliminary report on student achievement in college. Liberal education*. Washington, DC: AAC&U.
- Bailin, S., Case, R., Coombs, J. R., & Daniels, L. B. (1999). Common misconceptions of critical thinking. *Journal of Curriculum Studies*, 31(3), 269–283. doi:10.1080/002202799183124.
- Barnett, J. E., & Francis, A. L. (2012). Using higher order thinking questions to foster critical thinking: A classroom study. *Educational Psychology*, 32(2), 201–211.
- Barrow, R. (1991). The generic fallacy. *Educational Philosophy and Theory*, 23(1), 7–17. doi:10.1111/j.1469-5812.1991.tb00172.x.
- Bensley, D., & Spero, R. (2014). Improving critical thinking skills and metacognitive monitoring through direct infusion. *Thinking Skills and Creativity*, 12, 55–68. doi:10.1016/j.tsc.2014.02.001.
- Beyer, B. K. (2008). How to teach thinking skills in social studies and history. *The Social Studies*, 99(5), 196–201. doi:10.3200/TSSS.99.5.196-201.
- Brown, A. L. (1992). Design experiments. Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences*, 2(2), 141–178.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Davies, M. (2013). Critical thinking and the disciplines reconsidered. *Higher Education Research & Development*, 32(4), 529–544. doi:10.1080/07294360.2012.697878.
- Dwyer, C. P., Hogan, M. J., & Stewart, I. (2012). An evaluation of argument mapping as a method of enhancing critical thinking performance in e-learning environments. *Metacognition and Learning*, 7(3), 219–244. doi:10.1007/s11409-012-9092-1.
- Ennis, R. H. (1989). Critical thinking and subject specificity: Clarification and needed research. *Educational Researcher*, 18(3), 4–10. doi:10.3102/0013189X018003004.
- Ennis, R. H. (1993). Critical thinking assessment. *Theory into Practice*, 32, 179–186.
- Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London, UK: SAGE Publications Ltd..
- Garside, C. (1996). Look who's talking: A comparison of lecture and group discussion teaching strategies in developing critical thinking skills. *Communication Education*, 45, 212–227.
- Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist*, 39(2), 93–104. doi:10.1037/0003-066X.39.2.93.
- Halpern, D. F. (1993). Assessing the effectiveness of critical thinking instruction. *The Journal of General Education*, 42(4), 238–254.
- Halpern, D. F. (1998). Teaching critical thinking for transfer across domains. *American Psychologist*, 53(4), 449–455. doi:10.1037/0003-066X.53.4.449.
- Halpern, D. F. (2014). *Thought and knowledge: An introduction to critical thinking* (5th ed.). New York, NY: Psychology Press.
- Halpern, D. F. (2015). *Halpern critical thinking assessment*. Modling, Austria: Schuhfried GmbH.
- Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher*, 28(2), 16–25. doi:10.2307/1177186.
- Lawson, T. J. (1999). Assessing psychological critical thinking as a learning outcome for psychology majors. *Teaching of Psychology*, 26(3), 207–209.
- Lin, S.-S. (2014). Science and non-science undergraduate students' critical thinking and argumentation performance in reading a science news report. *International Journal of Science and Mathematics Education*, 12(5), 1023–1046. doi:10.1007/s10763-013-9451-7.
- McLean, C. P., & Miller, N. A. (2010). Changes in critical thinking skills following a course on science and pseudoscience: A quasi-experimental study. *Teaching of Psychology*, 37(2), 85–90. doi:10.1080/00986281003626714.
- McMillan, J. H. (1987). Enhancing college students' critical thinking: A review of studies. *Research in Higher Education*, 26(1), 3–29. doi:10.1007/BF00991931.
- McMurray, M. A. (1991). Reliability and construct validity of a measure of critical thinking skills in biology. *Journal of Research in Science Teaching*, 28(2), 183–192.
- McPeck, J. (1990a). Critical thinking and subject specificity: A reply to Ennis. *Educational Researcher*, 19(4), 10–12. doi:10.3102/0013189X019004010.
- McPeck, J. (1990b). *Teaching critical thinking: Dialogue and dialectic*. New York, NY: Routledge.
- Merrill, D. M. (2002). First principles of instruction. *Educational Technology Research and Development*, 50(3), 43–59.
- Merrill, D. M. (2013). *First principles of instruction: Identifying and designing effective, efficient, and engaging instruction*. San Francisco, CA: John Wiley & Sons, Inc..
- Moore, T. (2011). Critical thinking and disciplinary thinking: A continuing debate. *Higher Education Research & Development*, 30(3), 261–274. doi:10.1080/07294360.2010.501328.

- National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.
- Niu, L., Behar-Horenstein, L. S., & Garvan, C. W. (2013). Do instructional interventions influence college students' critical thinking skills? A meta-analysis. *Educational Research Review*, 9, 114–128. doi:[10.1016/j.edurev.2012.12.002](https://doi.org/10.1016/j.edurev.2012.12.002).
- Norris, S. P. (1989). Can we test validly for critical thinking? *Educational Researcher*, 18(9), 21–26. doi:[10.2307/1176715](https://doi.org/10.2307/1176715).
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.
- Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students: A third decade of research. (volume 2)*. San Francisco, CA: Jossey-Bass.
- Perkins, D., & Salomon, G. (1989). Are cognitive skills context-bound? *Educational Researcher*, 18(1), 16–25. doi:[10.3102/0013189X018001016](https://doi.org/10.3102/0013189X018001016).
- Resnick, L. (1987). *Education and training*. Washington, DC: National Academy Press.
- Resnick, L., Michaels, S., & O'Connor, M. (2010). How (well-structured) talk builds the mind. In D. D. Preiss & R. J. Sternberg (Eds.), *Innovations in educational psychology: Perspectives on learning, teaching, and human development* (pp. 163–194). New York, NY: Springer Publishing Company.
- Sendag, S., & Odabasi, H. (2009). Effects of an online problem based learning course on content knowledge acquisition and critical thinking skills. *Computers & Education*, 53(1), 132–141. doi:[10.1016/j.compedu.2009.01.008](https://doi.org/10.1016/j.compedu.2009.01.008).
- Siegel, H. (1988). *Educating reason: Rationality, critical thinking, and education*. New York, NY: Routledge.
- Smith, G. (2002). Are there domain-specific thinking skills? *Journal of Philosophy of Education*, 36(2), 207–227. doi:[10.1111/1467-9752.00270](https://doi.org/10.1111/1467-9752.00270).
- Solon, T. (2007). Generic critical thinking infusion and course content learning in introductory psychology. *Journal of Instructional Psychology*, 34(2), 95–109. Retrieved from <http://eric.ed.gov/ERICWebPortal/recordDetail?accno=EJ774169>.
- Spektor-Levy, O., Eylon, B.-S., & Scherz, Z. (2009). Teaching scientific communication skills in science studies: Does it make a difference? *International Journal of Science and Mathematics Education*, 7(5), 875–903. doi:[10.1007/s10763-009-9150-6](https://doi.org/10.1007/s10763-009-9150-6).
- Stark, E. (2012). Enhancing and assessing critical thinking in a psychological research methods course. *Teaching of Psychology*, 39(2), 107–112. doi:[10.1177/0098628312437725](https://doi.org/10.1177/0098628312437725).
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson Education, Inc..
- Tiruneh, D. T., De Cock, M., Weldeslassie, A., Elen, J., & Janssen, R. (2017). Measuring critical thinking in physics: Development and validation of a critical thinking test in electricity and magnetism. *International Journal of Science and Mathematics Education*, 15(4), 663–682.
- Tiruneh, D. T., Verburgh, A., & Elen, J. (2014). Effectiveness of critical thinking instruction in higher education: A systematic review of intervention studies. *Higher Education Studies*, 4(1), 1–17. doi:[10.5539/hes.v4n1p1](https://doi.org/10.5539/hes.v4n1p1)
- Toy, B. Y., & Ok, A. (2012). Incorporating critical thinking in the pedagogical content of a teacher education programme: Does it make a difference? *European Journal of Teacher Education*, 35(1), 39–56. doi:[10.1080/02619768.2011.634902](https://doi.org/10.1080/02619768.2011.634902).
- Tsui, L. (1999). Courses and instruction affecting critical thinking. *Research in Higher Education*, 40(2), 185–200.
- van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: A four-component instructional design model for teaching technical training*. Englewood Cliffs, NJ: Educational Technology.
- Wheeler, L. A., & Collins, S. K. R. (2003). The influence of concept mapping on critical thinking in baccalaureate nursing students. *Journal of Professional Nursing*, 19(6), 339–346. doi:[10.1016/S8755-7223\(03\)00134-0](https://doi.org/10.1016/S8755-7223(03)00134-0).
- Williams, R. L., Oliver, R., & Stockdale, S. L. (2004). Psychological versus generic critical thinking as predictors and outcome measures in a large undergraduate human development course. *The Journal of General Education*, 53(1), 37–58. doi:[10.1353/jge.2004.0022](https://doi.org/10.1353/jge.2004.0022).